iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** VI **Month of publication:** June 2020

DOI: http://doi.org/10.22214/ijraset.2020.6233

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VI June 2020- Available at www.ijraset.com

Credit Card Fraud Detection using Machine Learning Methods

Sanath Kumar Bhat K[1], Sreenath N[2], Divya B S[3]
[1, 2]*Dept. of Information Science and Engineering, Visvesvaraya Technological University Sapthagiri College of Engineering Bangalore, India*
[3]*Assistant professor, Dept. of Information Science and Engineering Visvesvaraya Technological University Sapthagiri College of Engineering Bangalore, India*

Abstract: A major problem which is affecting growth in financial services is "CREDIT CARD FRAUD". Many organizations lost their amount due to these frauds. Even though many research studies are made on fraud detection, they lack on analyzing data extracted from actual transaction, due to privacy issues. Here, certain machine learning algorithms are used to detect fraudulent transactions.
First, the Standard methods are applied. Then, in combination hybrid methods such as AdaBoost and majority voting are applied. A publicly available datasets are used so that model efficacy can be evaluated. Later, a real-world credit card data set is analyzed which is taken from certain financial institution. Further, to estimate the robustness of the algorithm, noise is added to the samples.
By comparing various machine learning algorithms, the main aim is to find the best in those to detect the fraudulent transactions to avoid credit card fraud. The experimental results indicate that the hybrid methods such as majority voting efficiently provides nearly best accuracy for detecting fraudulent transactions of Credit cards.
Keywords: AdaBoost, classification, credit card, accuracy, MCC Score, majority voting, fraud transaction prediction.

I. INTRODUCTION

In technical terms, criminal deception which brings personal or financial gain is stated as fraud. To avoid this loss, we can follow two mechanisms i.e. fraud prevention & fraud detection. As we say "prevention is better than cure" People has to be careful and follow preventive measures that involves proactive method i.e. Fraud prevention which prevents the fraud to happen in the first place. Other option to recur the loss of fraudulent transaction is fraud detection.

The illicit use of credit card or its information is considered as Credit card fraud. The transactions using credit card can be classified in two ways i.e. usage of credit cards itself in person and in digital transactions fraudsters use the confidential information that includes credit number, expiry date & OTP i.e. verification number to accomplish transaction through internet or telephone.

The rapid increase of digitalized payments includes usage of credit cards. The statistics of Reserve bank of India states that they are issuing approximately million new credit cards yearly which shows the rise of usage in credit cards in India The value of total transaction of credit cards in 2018-19 is six lakh crore i.e. 30% higher from the 4.6 lakh crore in 2017-18.This growth is affected by frauds which has been constantly increased even though high level security includes many authorization techniques.

The financial or personal loss affects directly the merchants and including all costs like card issuer fees and charges. These losses has to be tolerated by merchants so in order to balance the goods are rated higher.

To diminish this loss, an operative system of detecting fraud is necessary to lessen or eradicate this fraud cases. Various studies have been made on detecting fraudulent transactions of credit card. Finally some methods of machine learning like "artificial neural networks, rule- induction techniques, decision trees, logistic regression, and support vector machines" are useful. Several methods can be used in combined manner or standalone

Nearly twelve machine learning algorithms which are ranging from standard neural networks to deep learning methods are suggested in this paper to detect credit card fraud and evaluation is done by real world and benchmark datasets. Hybrid model is formed by AdaBoost and popular voting methods.

Later to the datasets, noise is added to assess the reliability and robustness of the models. Evaluating the different methods of machine learning with the actual credit card datasets is the key contribution of this paper. The datasets extracted from actual credit card transaction has been used here.

This paper is organized is as follows, Section I has the introduction part. Section II includes related previous studies and researches conducted. Section III includes various algorithms of machine learning which is approached. Section IV has the experiments with bench mark & real-world credit card datasets. Section V includes conclusion and future enhancement.

II. LITERATURE SURVEY

Y Sahin et al [1], have approached new cost-sensitive decision tree algorithm for fraud detection. To measure the performance, splitting attribute at each non-terminal node is selected by reducing the sum of misclassification costs & comparing it with familiar classification model with real world credit card data set. In this study, Varied misclassification costs are shown. When these experimental results of cost-sensitive algorithm are compared with existing familiar methods the performance is much better with respect to accuracy, positive rate metrics and also defines a cost-sensitive metric specific to credit fraud detection. Fraudulent transactions can be avoided by this approach to reduce financial losses.
Demerit: Inaccuracy in Fraud Detection

A.O.Adewumi et al [2] have submitted a work which includes survey of nature based and machine learning fraud detection techniques and concentrated on difficulties in detection of fraud in credit card transactions. Merits and demerits of different machine learning methods are identified and compared. The mentioned techniques are classified as misuses(supervised) & anomaly detection (unsupervised). Another classification is based on ability to process numerical & categorical datasets. For further use, common attributes are extracted from real world and datasets which are used in study are described and organized into real and synthesized data. Mainly criteria required for evaluation of techniques are also discussed.
Demerit: Accuracy is too Low

A.Srivastava et al [3] proposed (HMM) Hidden Markov Model to model the sequence of operations credit card transactions & how it is used for fraud detection. The normal behavior of card holder i.e pattern is recorded. Incoming credit card transaction is cancelled if trained HMM rejects with high probability and it is classified as fraudulent and it is also important to safeguard the genuine credit card transaction such that to take care it is not rejected.
Demerit: For Wide Variation Of Input Data, Accuracy Is Only Close To 80%

J.T Quah et al [4] have submitted a work which focuses on real-time fraud detection based on computational intelligence. This innovative approach is used to understand spending patterns in order to decipher potential fraud cases i.e. Self- Organization Map is used in deciphering, filtering and analyzing pattern behavior of customer for detection of fraud.
Demerit: Accuracy Can Be Improvised In Better Way With More Efficient Models

A. Mishra et al [5] have approached various classification and grouping techniques for detection of fraud cases in credit card. It states that the possibility of the fraud transaction is very less but not negligible. So their work aims at judging various classifiers by inspecting various classification techniques. It focuses on improving fraud detection and doesn't misclassify genuine transaction as fraud.
Demerit: Accuracy is low when compared to other models.

III. PROPOSED SYSTEM

In this research, to detect fraudulent credit card transactions, various machine learning algorithms are used. In combination, we are applying AdaBoost and majority voting algorithms in order to form the hybrid models. So, the system is fast because of AdaBoost and Majority voting algorithm.
In our proposed system user has to upload the dataset in the beginning and then split the data into train data and test data in order to train the model. Once model gets trained starts testing based on the given inputs. Finally the model predicts the transaction as fraud by using hybrid algorithms such as AdaBoost and majority voting.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VI June 2020- Available at www.ijraset.com

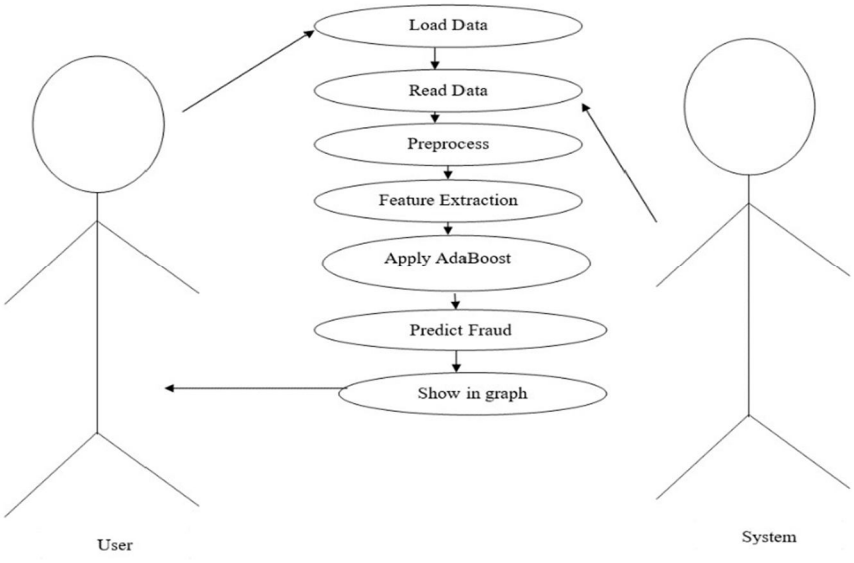


Figure 1: Use case diagram

A. Advantages of Proposed System

1) High Accuracy in Detection of fraud transaction
2) Less time for processing large data
3) Less resource consumption
4) Energy Efficient
5) Efficient classification of sub classes in identifying fraud transaction.

IV. METHODOLOGY

Algorithms
Two major algorithms are used in this experimental study:-

A. ADABOOST

AdaBoost algorithm is ensemble boosting classifier proposed in 1996 by Robert Schapire and Yoav Freund. In this algorithm multiple classifiers are combined to boost the accuracy. A strong classifier is build by AdaBoost by combining poorly performing multiple classifiers finally, we will get classifier having high accuracy. The idea behind this AdaBoost is to set the classifiers weights and to train the sample data in every iteration such that it provides accurate predictions. Machine learning algorithm which is accepting weights on the training set is used as base classifier. AdaBoost has to meet the following two conditions:

1) Training examples with different weighed has to be iteractively trained by the classifier.
2) In order to provide an excellent fit for examples, it minimizes the training error for every iteration.

It works in the following steps:

a) The training subsets are selected by AdaBoost in the beginning..
b) AdaBoost model is iteratively trained by choosing the train set depending on the last training's accurate prediction.
c) To the wrongly classified observations, high weight is assigned so that in next iteration wrong classified observations will get the classification with high probability.
d) It also assigns the weight to the trained classifier in every iteration according to the classifier accuracy. The classifier will gets high weight if it is having more accuracy.
e) This process will iterate until it reach the specified maximum number of estimators or until the complete training data fits without any error
f) In order to classify, perform a "vote" across all of the model gets trained starts testing based on the given inputs. learning algorithms which was built.

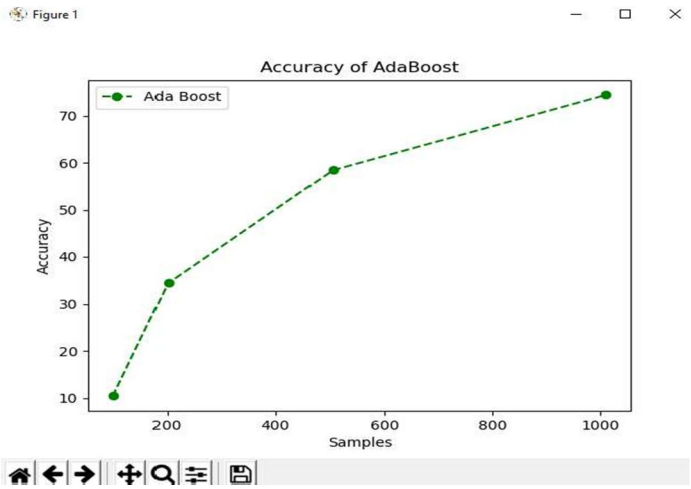


Figure 2: Accuracy graph of AdaBoost

B. Majority Voting

Majority voting is a hybrid algorithm is used in data classification. It is used for combining two or more algorithms in order to increase the accuracy of the model, prediction is made by each algorithm for every test sample. The majority of votes is considered as final output.

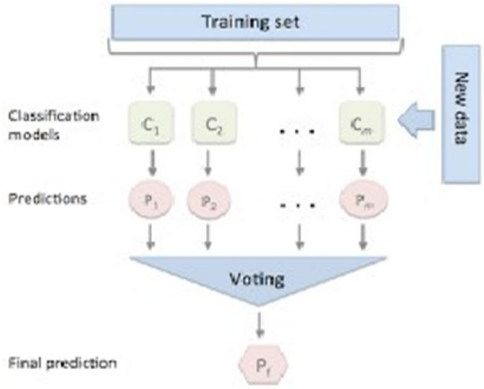


Figure 3:ajority voting

Here, we predict the class label y^ via majority (plurality) voting of each classifier Cj:

$$\hat{y} = mode\{C_1(\mathbf{x}), C_2(\mathbf{x}), \ldots, C_m(\mathbf{x})\}$$

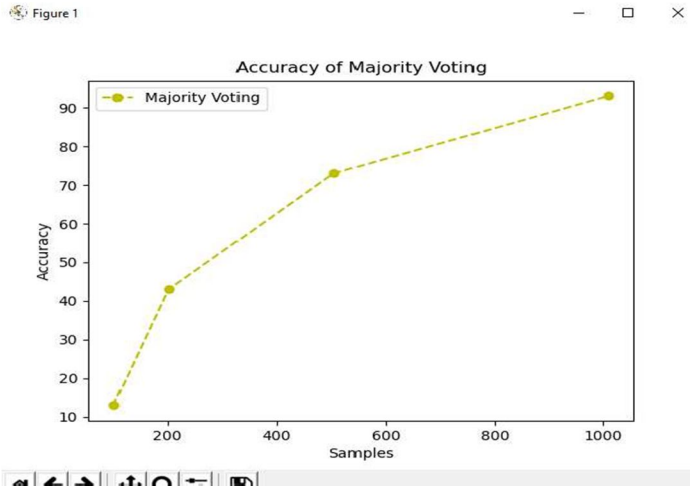


Figure 4: Accuracy graph of Majority voting

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VI June 2020- Available at www.ijraset.com

V. FUTURE ENHANCEMENTS

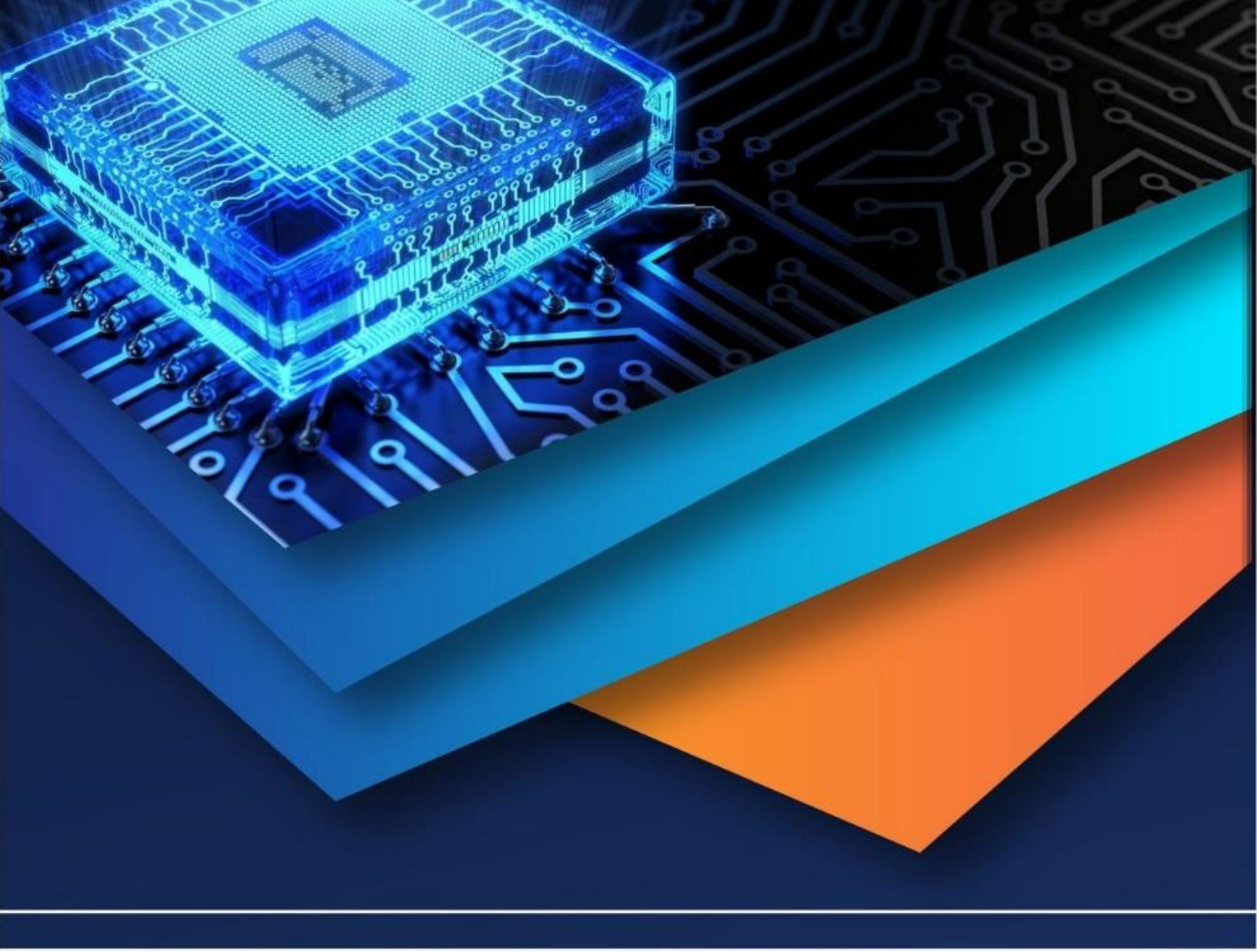
For future work, online training models can be adopted by studying the methods used in this paper. Also, investigation of the other training models can be done. The fast tracking of fraud cases can be achieved by online training models. This will help to detect fraud credit card transactions so that it can be prevented before they take place. Thus, in financial sectors, the losses incurred are reduced.

VI.CONCLUSION

In this paper, detection of fraudulent transactions in credit card has been studied using machine learning. First, many standard models are used such as NB, SVM and DL in practical evaluation. To evaluate, real-world credit card datasets have been used. Along with standard models, some hybrid models such as AdaBoost and majority voting have been used. To measure performance, the MCC metric have been adopted so that outcomes predicted as the true for positive and false for negative outcomes. "Majority voting gives the best MCC score as 0.823". From a financial institution, a real credit card datasets have also been used to evaluate. The same standard and hybrid models have been used. By using AdaBoost and Majority voting a perfect MCC score of 1 will be attained. To the data samples a noise from 10% to 30% has also been added to further estimate the hybrid models. "The best MCC score of 0.942 for 30% of noise added to the datasets have been achieved by majority voting". Thus, it shows that even in the presence of noise, robust performance is offered by majority voting method.

REFERENCES

[1] Y.Sahin, S.Bulkan and E.Duman(2013)-"A cost-sensitive decision tree approach for fraud detection"
[2] A. O. Adewumi and A. A. Akinyelu(2016)-"A survey of machine-learning and nature-inspired based credit card fraud detection techniques"
[3] A.Srivastava,A.Kundu,S.Sural,andA.Majumdar(2008) - "Credit card fraud detection using hidden Markov model"
[4] J. T. Quah and M. Sriganesh(2007)-"Real-time credit card fraud detection using computational intelligence"
[5] A.Mishra,C.Ghorpade(2018)-"Credit Card Fraud Detection on the Skewed Data Using Various Classification and Ensemble Techniques"
[6] S. Lakshmi, S. D. Kavilla(2018)-"Machine Learning For Credit Card Fraud Detection System"
[7] N. Malini, Dr. M. Pushpa(2017)-" on Credit Card Fraud Identification Techniques based on KNN and Outlier Detection"
[8] Z. Kazemi, H. Zarrabi(2017)-"Using deep networks for fraud detection in the credit card transactions"
[9] C. Wang, Y. Wang, Z. Ye, L. Yan, W. Cai, S. Pan(2019)- "Credit card fraud detection based on whale algorithm optimized BP neural network"
[10] N. Kalaiselvi, S. Rajalakshmi, J. Padmavathi(2018)-"Credit card fraud detection using learning to rank approach"



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY